QMRF identifier (JRC Inventory):	
QMRF Title: *QSAR model for Persistence - Abiotic degradation in air, OH Tropospheric Degradation of Volatile Organic Compounds*	
Printing Date: *18.05.2010*	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Persistence - Abiotic degradation in air, OH Tropospheric Degradation of Volatile Organic Compounds

1.2.Other related models:

1.3.Software coding the model:

QSARModel 4.0.4 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

08.02.2010

2.2.QMRF author(s) and contact details:

[1]Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[2]Tarmo Tamm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[3]Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[4]Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[5]Dana Martin Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[6]Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[7]Deniss Savchenko Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[8]Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[9]Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[10]Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[11]Mati Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com

2.6.Date of model development and/or publication:

07.02.2010

2.7.Reference(s) to main scientific papers and/or software package:

2.8.Availability of information about the model:

All information in full detail is available

2.9.Availability of another QMRF for exactly the same model:

None to date

3.Defining the endpoint - OECD Principle 1

3.1.Species:

n/a

3.2.Endpoint:

2.Environmental fate parameters 2.Persistence: Abiotic degradation in air (Phototransformation) 2.2.b.Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)

3.3.Comment on endpoint:

Rate constant for OH radical degradation.

The dominant chemical process of chemicals in the gasphase is their reaction with OH radicals, NO3 radicals, and ozone. The hydroxyl radical is the key reactive species in the troposphere, where it reacts with practically every organic compound.

3.4.Endpoint units:

cm3 s-1 molecule-1

3.5.Dependent variable:

-logK(OH) (original rate constants were transformed into log scale and multiplied by -1 to reduce data range and obtain positive values)

3.6.Experimental protocol:

The selected data are for reactions at 25 °C and 1 atm. The gas-phase reaction rate constants of OH radical and organic chemicals have been directly measured.

3.7.Endpoint data quality and variability:

Original experimental data was collected from ref 3.

Statistics for -logK(OH):
max value: 15.7
min value: 9.44
standard deviation: 1.03
skewness: 1.16

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

2D and 3D regression-based QSAR

4.2.Explicit algorithm:

multilinear regression QSAR

multilinear regression QSAR derived with BMLR (Best Multiple Linear Regression) method

-logK(OH) = 3.61
+2.15*HASA-1/TMSA (AM1) (all)
-0.698*HOMO energy (AM1)
+1.67*Relative number of aromatic bonds
-12.7*HACA-1/TMSA (Zefirov)

4.3.Descriptors in the model:

[1]HASA-1/TMSA (AM1) (all) [unitless] relative solvent-accessible surface area of H-bonding acceptor atoms (from AM1 calculation)

[2]HOMO energy (AM1) [eV] energy of highest occupied molecular orbital energy

[3]Relative number of aromatic bonds [unitless] Relative number of aromatic bonds

[4]HACA-1/TMSA (Zefirov) [unitless] sum of solvent-accessible surface area of H-bonding acceptor atoms, selected by threshold charge

4.4.Descriptor selection:

Initial pool of ~1000 descriptors for each structure calculated. Stepwise descriptor selection was applied to reduce the pool based on a set of statistical selection rules (for one-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold), two parameter correlations developed from previously reduced pool, the statistical selection applied: (intercorrelation coefficient below threshold, significant correlation with endpoint, in terms of correlation coefficient and t-test)

Stepwise trial of additional descriptors not significantly correlated to any already in the model.

See refs 1-2.

4.5.Algorithm and descriptor generation:

1D, 2D, and 3D theoretical calculations. Descriptors derived from mol files. Quantum chemical descriptors from AM1 calculations. Model developed by using multilinear regression using ordinary least squares.

4.6.Software name and version for descriptor generation:

QSARModel 4.0.4

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

53 (212 chemicals / 4 descriptors)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Applicability domain based on training set:

a) by chemical identity: Diverse set of Volatile Organic Compounds (alphatic and aromatic hydrocarbons, alcohols, amines, halogenated compounds, etc)

b) by descriptor value range: The model is suitable for compounds that have the descriptors

in the following minimal-maximal range:

HASA-1/TMSA (AM1) (all): 0 - 0.911

HOMO energy (AM1): -13.3 - -8.10

Relative number of aromatic bonds: 0 - 0.615

HACA-1/TMSA (Zefirov): 0 - 0.0587

5.2.Method used to assess the applicability domain:

By chemical identity - compounds must be similar to traing set compounds in terms of functionality.

By descriptor value range: range of descriptor values similar to training set with ±30% confidence. Descriptor values must fall between maximal and minimal descriptor values of training set ±30%.

5.3.Software name and version for applicability domain assessment:

QSARModel 4.0.4

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

5.4.Limits of applicability:

See 5.1

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN:Yes

Chemical Name:Yes

Smiles:No

Formula:Yes

INChI:No

MOL file:Yes

6.3.Data for each descriptor variable for the training set:

All

6.4.Data for the dependent variable for the training set:

All

6.5.Other information about the training set:

212 data points (Original source dataset split into testing and training. From the original source data of 423 values sorted by endpoint value, each 2nd was subjected to the test set.)

0 negative values
212 positive values

6.6.Pre-processing of data before modelling:

no more than specified in 3.5

6.7.Statistics for goodness-of-fit:

R2 = 0.832 (Correlation coefficient)
s2 = 0.427 (Standard error of the estimate)
F = 256.8 (Fisher function)

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

R2CV = 0.821

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

R2CVMO = 0.819 (80% : 20% , training : testing)

6.10.Robustness - Statistics obtained by Y-scrambling:

n/a

6.11.Robustness - Statistics obtained by bootstrap:

n/a

6.12.Robustness - Statistics obtained by other methods:

ABC analysis (2:1 training : prediction) on sorted (in increased order of endpoint value) data divided into 3 subsets (A;B;C). Training set formed with 2/3 of the compounds (set A+B, A+C, B+C) and validation set consisted of 1/3 of the compounds (C, B, A).

average R2 (fitting) = 0.833
average R2 (prediction) = 0.824

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

Yes

7.2.Available information for the external validation set:

CAS RN:Yes
Chemical Name:Yes
Smiles:No
Formula:Yes
INChI:No
MOL file:Yes

7.3.Data for each descriptor variable for the external validation set:

All

7.4.Data for the dependent variable for the external validation set:

All

7.5.Other information about the external validation set:

211 data points
0 negative values,
211 positive values

7.6.Experimental design of test set:

Original source dataset split into testing and training. From the original source data, sorted by endpoint value, each 2nd was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.773 (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

All are in range of applicability domain:

HASA-1/TMSA (AM1) (all): 0 - 0.942

HOMO energy (AM1): -13.1 - -8.27

Relative number of aromatic bonds: 0 - 0.579

HACA-1/TMSA (Zefirov): 0 - 0.0561

7.9.Comments on the external validation of the model:

The validation coefficient of determination (R2) is signifficant and close to those coefficients of internal validation (R2CV and R2CVMO).

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Descriptors "HASA-1/TMSA (AM1) (all)" and "HACA-1/TMSA (Zefirov)" are simultaneously taking into account a capability of hydrogen acceptor bonding and the size of the compound. Although the descriptors seem to be similar, they are counting different features of hydrogen acceptor abilities. "HASA-1/TMSA (AM1) (all)" counts all possible hydrogen acceptor atoms solvent accessible surface area while "HACA-1/TMSA (Zefirov)" counts only charged areas. "HOMO energy (AM1)" is an indicator of the nucleophilicity of the molecule - reactive molecules have relatively higher HOMO energy.. "Relative number of aromatic bonds" represents a count of aromaticity which differentiates these compounds from aliphatic ones. All presented descriptors are representing important molecular properties related to H-abstraction.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments (in ref. 4. HOMO energy and aromatic carbond were found to be important,

8.3.Other information about the mechanistic interpretation:

Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training?Test Sets and Consensus Modeling

Paola Gramatica, Pamela Pilutti, and Ester Papa

J. Chem. Inf. Comput. Sci., 2004, 44 (5), pp 1794–1802

9.Miscellaneous information

9.1.Comments:

Similar methodology to current approach was used in refs 1-2

9.2.Bibliography:

[1]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC 16, 38-60. http://dx.doi.org/10.1021/ci049923u

[2]Karelson M, Karelson G, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A,

Savchenko D & Dobchev D (2009). QSAR study of pharmacological permeabilities. ARKIVOC 2, 218–238.
[3]Atkinson, R. Kinetics and mechanisms of the gas-phase reactions of the hydroxyl radical with organic compounds. Journal of Physical Reference Data, Monograph 1, 1989.
[4]Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training?Test Sets and Consensus Modeling Paola Gramatica, Pamela Pilutti, and Ester Papa J. Chem. Inf. Comput. Sci., 2004, 44 (5), pp 1794–1802

9.3.Supporting information:

Training set(s)Test set(s)Supporting information

Karelson Arkivoc 2008	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2008.pdf
Karelson Arkivoc 2009	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2009.pdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

10.4.Comments: